

An Optimized Feature Matching Method Based on Rolling Guided Filter Constrains

Xiaomin Liu¹, Yizhao Yuan¹, Donghua Yuan¹, Jeng-Shyang Pan², Huaqi Zhao^{1*}

¹ The Heilongjiang Provincial Key Laboratory of Autonomous Intelligence and Information Processing, Jiamusi University, China

² College of Computer Science and Engineering, Shandong University of Science and Technology, China
xiaominliu@vip.sina.com, zzwylx@163.com, 674861335@qq.com, jsyan@ieee.org, zhaohuaqi@126.com

Abstract

Unmanned aerial vehicle (UAV) cross-view object localization is widely used in applications such as field rescue, disaster detection, and express delivery. The feature-matching method, which locates the object position by matching the UAV object image and satellite image, is a key technology for UAV cross-view object localization. This study proposes a feature-matching method optimized using a rolling-guided filter. First, a rolling-guided filter is introduced to calculate the keypoint correspondence, which addresses the problem of scale difference between the two images; then, we introduce the phase congruency information for feature description. By obtaining candidate corresponding keypoint sets through feature matching, the problem of gray level differences between images can be solved; Finally, we use the correspondence of candidate keypoints to calculate the network loss, which addresses the problem of viewing angle differences between images. The experimental results demonstrate that the proposed method can effectively improve the feature-matching performance.

Keywords: Image matching, Point matching, Set matching, Diffeomorphic, Phase congruency

1 Introduction

Unmanned aerial vehicles (UAVs) have become a research hotspot in many countries because of their small size, high flexibility, and low cost. Currently, drones are being widely used in field rescues [1], disaster detection [2], precision agriculture [3], surveying and mapping [4], and express delivery [5]. Most UAV applications require precise localization of objects; therefore, object-localization technology is the basis for many applications. Object localization method based on feature matching is an autonomous positioning technique. In this method, the given satellite reference image is matched to the object image captured by a UAV, which is also known as cross-view object localization. As the satellite reference image and UAV object image contain an obvious differences, this method is also called cross-view object localization [6-9].

With the development of computer vision, several methods have been proposed by scholars. Zhang et al. proposed a 3D reconstruction algorithm for small indoor scenes, which improved the matching algorithm and matching strategy, and advanced the real-time property of motion recovery algorithm [10]. Duan et al. proposed a UAV image stitching algorithm based on image segmentation and generalized Hough transform based on the existing image stitching algorithm. [11]. Li et al. accomplished large-scale intelligent driving scene reconstruction by combining stereo matching, depth-value calculation, triangulation, and texture mapping algorithms [12]. Liu et al. proposed the construction and matching of thermal feature description method based on image retrieval, that solved the problem of mismatches caused by highly similar local descriptors in different regions during the feature matching process [13]. Zou et al. proposed a fully automatic stitching method for UAV images based on 3D reconstruction to obtain geometrically consistent stitching images [14]. However, these methods cannot solve the problems of viewing angle differences and large-scale deformations between satellite reference and object images. To address these two problems, this study analyses the research status from three perspectives: Region-, feature-, and learning-based matching.

The region-based matching method defines a similarity measure criterion (such as a normalized cross correlation [15], mutual information [16], and cross cumulative residual entropy [17], etc.), the similarity is calculated using the sliding window method according to the comparison of current window and the template image, and the matching is successful when the similarity is the largest. Wu proposed a fast NCC matching method based on wavelet pyramid search strategy [18]. Yu et al. proposed a secondary matching method based on image edge features [19]. Fu proposed a fast anti-rotation matching algorithm based on covariance matrix [20]. Owing to the need to traverse all possible transformation relationships between images, the calculation of the region-based matching method is time-consuming. Therefore, in case of large gray-scale differences, scale changes, and rotation changes between images, the matching is likely to fail.

Feature-based matching is the algorithm most widely used for this purpose. This method first extracts salient features of the image (such as point feature [21-22], line

*Corresponding Author: Huaqi Zhao; Email: zhaohuaqi@126.com
DOI: <https://doi.org/10.70003/160792642025122607002>

feature [23], surface feature [24] and optimizing feature [25], and then determines the matching feature point pair by calculating the distance of the features in the two images, and then finally selects the appropriate spatial transformation model. The traditional scale invariant feature transform (SIFT) algorithm is proposed by Lowe et al. [26]. Ke et al. combined the principal components analysis (PCA) method with the SIFT algorithm to improve its efficiency [27]. Bay et al. proposed Speeded Up Robust features (SURF) and Kova et al. applied it to finger vein recognition, which uses box filters to approximate the differential Gaussian of SIFT and performs feature detection on a fast Hessian matrix [28]. Gao et al. proposed an optimized affine invariant SIFT algorithm [29]. Wang et al. applied the SURF algorithm to match UAV images with the satellite images [25]. Niu et al. proposed a fast feature matching method based on a scale-invariant feature transformation [30]. Zhang et al. proposed a UAV feature matching algorithm based on CenSurE-star [31]. The feature-based matching method has a high computational efficiency, and good anti-occlusion and anti-geometric distortion. However, no method can perfectly solve this complex problem of matching satellite reference images and UAV object images for UAV cross-view object localization.

With the rise of artificial intelligence, feature matching methods based on deep learning have achieved significant progress. Yi et al. proposed a learning-based invariant feature transformation called LIFT [32] which is a pioneering method that implements a complete feature matching process. Yuan et al. used VGG16 convolution features as keypoint descriptors to improve the matching performance between UAV and satellite images [33]. A part of this study also aims at obtaining a feature extraction model by training a deep convolutional neural network (CNN) [34-36]. Choi et al. segmented an image through a neural network to obtain the building area and circular area and used the ratio of the building area to a circular area as a feature for complete matching [37]. Nassar proposed deep CNN-based registration boxes for the UAV images [38]. Xu et al. proposed a joint description network [39]. These methods are more effective than traditional feature-based matching methods; however, it required a large amount of data, which limits its application to UAV cross-view object localization.

According to the above analysis, the following problems exist in UAV cross-view object localization. First, traditional approach cannot solve the problems of viewing angle difference and large-scale deformation between the satellite reference image and the object image. Second, the traditional feature matching method is prone to failure when large grayscale differences exist. Last, the state-of-the-art deep learning-based feature matching methods only consider certain attributes such as sparsity, repeatability, and distinctiveness, which may limit accuracy in a complex circumstance with noise.

To address these problems, this study focuses on feature matching methods, combining deep learning-based feature matching methods to propose a rolling-guided filter-based optimized feature matching method and make

the following contributions:

(1) A rolling-guided filter was introduced to generate an image-scale pyramid that preserved the image edge information and improved the repeatability of keypoints.

(2) A phase-weighted information was introduced to construct feature descriptors that improved the adaptability of feature descriptors to large grayscale differences between images. The above method was then used to obtain the candidate corresponding keypoints of an image. The corresponding keypoint is employed to compute the network loss, which introduces attribute constraints to train the keypoint detector and feature descriptor.

(3) We introduced the corresponding keypoint set obtained by nearest neighbour matching to better completing the feature-matching task between the satellite reference and UAV object images.

2 Related Works

Feature matching methods are oriented toward UAV crossview object localization tasks, and can be classified as region-based and feature-based methods for different matching information. The region-based matching method compares the similarity between the windows using the sliding window approach and the two images are considered as successfully matched when the similarity is optimal. The feature-based methods include point, line, and surface features. Point features are the most commonly used and researched features, which generally exist at locations where the gray value changes significantly in all directions in the image. Point features have a better resistance to occlusion and are easy to extract as well as more suitable for matching satellite images and UAV images than face features and line features. Feature-matching methods can be further divided into traditional and learning-based feature-matching methods. Based on the traditional feature matching methods, this study proposed a feature-matching method optimized using a rolling-guide filter, and introduced phase congruency information to create a feature descriptor. The three methods are described below.

2.1 SIFT: A Feature Matching Approach

The targets of early point-feature matching research were mainly the corner points of the image. Harris et al [40] proposed the corner point extraction algorithm in 1988, using differential operations and an autocorrelation matrix to detect corner points in an image with a rotation invariance and illumination invariance, but without scale invariance. Mikolajczyk et al [41] improved the Harris corner point extraction algorithm. Lowe proposed the SIFT algorithm, which uses a differential Gaussian operator to approximate the Laplacian operator to establish the differential scale space of the image. SIFT is a classical feature-matching method with good performance in natural image matching. SIFT feature matching algorithm comprises five main steps: scale-space extremum detection, keypoint location, assignment of principal directions, construction of feature descriptors, and feature

matching. The scale space L of the image was obtained from the following equation:

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y) \quad (1)$$

where $G(x, y, \sigma)$ is a variable-scale Gaussian function, $I(x, y)$ is input image, the symbol $*$ is convolution operator. The difference in the Gaussian (DoG) image space is generated by convolving an image with Gaussian kernels of adjacent scales as shown in the following equation:

$$D(x, y, \sigma) = (G(x, y, k\sigma) - G(x, y, \sigma)) * I(x, y) \quad (2)$$

The relationship between the difference in the Gaussian (DoG) operator and the normalized Laplacian operator is as follows:

$$G(x, y, k\sigma) - G(x, y, \sigma) \simeq (k - 1)\sigma^2 \Delta^2 G \quad (3)$$

Therefore, using the DoG operator instead of the LoG operator can improve the keypoint detection efficiency [40].

The rolling-guided filter is an edge-preserving filter that can retain the edge information in images, which is beneficial for improving the repeatability of keypoints. Therefore, in the subsequent sections of this study, a rolling-guided filter was used instead of a Gaussian filter to generate the scale pyramid.

2.2 Phase Congruency Feature Detection Method

Phase congruency-based feature detection methods utilize the moments of phase congruency information to determine the corners and edges. Phase congruency is a dimensionless quantity that provides contrast invariant information; thus, the magnitude of the main moments of phase congruency can be used directly to determine the strength of the corners and edges [42]. Kovess improved the phase congruency model and successfully applied it to feature detection using the following model structure:

$$PC_2(x) = \frac{\sum_n W(x) B_n(x)}{\sum_n A_n(x) + \epsilon} \quad (4)$$

$$B_n(x) = [A_n(x)(C(x) - |S(x)|) - T] \quad (5)$$

$$C(x) = \cos(\phi_n(x) - \bar{\phi}(x)) \quad (6)$$

$$S(x) = \sin(\phi_n(x) - \bar{\phi}(x)) \quad (7)$$

where function $W(x)$ represents the weight factor for the frequency extension. To avoid division by zero,

a small constant value ϵ is added. Only when the energy response is greater than T is included in the calculations. The notation $[p]$ indicates that p is equal to itself when it is positive; otherwise, it is zero.

According to the classical moment analysis equation, the maximum and minimum moments of phase congruency are denoted by M and m respectively,

$$M = \frac{1}{2} (c + a + \sqrt{b^2 + (a - c)^2}) \quad (8)$$

$$m = \frac{1}{2} (c + a - \sqrt{b^2 + (a - c)^2}) \quad (9)$$

where M represents the edges in the image and m represents the corners in the image. And a , b , and c are calculated as follows:

$$a = \sum (PC(\theta) \cos(\theta))^2 \quad (10)$$

$$b = 2 \sum (PC(\theta) \cos(\theta)) \cdot (PC(\theta) \sin(\theta)) \quad (11)$$

$$c = \sum (PC(\theta) \sin(\theta))^2 \quad (12)$$

2.3 Feature Matching based on Property Optimization

For convenience, we first provide the meanings of some of the symbols. Let F and D denote the keypoint detector and feature descriptor, respectively. The output score of the keypoint detector is defined as:

$$f_j[i] = nms_{rad}(F(o_i, T_j(I)|\theta_F)), i \in 1, \dots, N \quad (13)$$

where nms_{rad} denotes the non-maximum suppression function that ensures that one keypoint can be detected in a local window with a radius of rad . I represents the original image in the training sample and N is the number of pixels in the image. Under illumination and viewpoint variations, I is transformed into $T_j(I)$, where j denotes the different transformations. o_i represents the corresponding pixel in the original image I and transformed image $T_j(I)$.

Descriptor D outputs a descriptor vector for each keypoint in the image $T_j(I)$, which is defined as follows:

$$d_j[i] = D(o_i, T_j(I)|\theta_F), i \in 1, \dots, N \quad (14)$$

where $d_j(i)$ is a normalized unit vector.

The objective of the training process is to maximize the joint probability of all attributes to optimize the keypoint detector and feature descriptor [43]. As the locations of the keypoints in the image are not known in advance, a latent variable y is introduced to represent a set of keypoints. Y is a vector of length N (number of pixels in the image), where

$y[i] = 1$ if pixel o_i is a keypoint, or $y[i] = 0$. According to the definition of o_i , $y[i] = 1$ implies that o_i is a keypoint in both the original image I and the transformed image $T_j(I)$.

Using the attribute optimization method described above, we can theoretically obtain keypoints with repeatability and distinctiveness. However, the performance of the cross-view feature matching scenario decreased.

3 Proposed Method

3.1 Feature Matching Architecture based on Rolling Guided Filter

The overall framework of the method is illustrated in Figure 1. First, the satellite reference image and UAV object image are input into the framework. Second, a rolling-guided filter is used to establish a difference-of-Gaussian rolling-guided scale space (DoGR). Next, phase weighted information is used to construct the feature descriptors for each keypoint, and candidate corresponding keypoints are obtained through nearest-neighbour matching. The response of the corresponding candidate keypoints are used as constraints to construct a network loss. Finally, the proposed model is used to match the satellite reference and UAV object images. Thus, this framework includes the following three optimizations:

(1) This study proposes a method of computing keypoint response based on rolling-guided filter. To address the issue of losing image edge information, a rolling guided filter is used to build an image scale pyramid, which retains the edge information of the image.

(2) A feature description based on phase-weighted information is explored to describe the feature. Phase congruency information is introduced to construct a feature descriptor based on phase-weighted information. Candidate corresponding keypoint sets are obtained through feature matching using this descriptor.

(3) A property optimization based on keypoint response constraint is conducted to accomplish image matching in this study. A network loss constraint is constructed using the keypoint responses of corresponding keypoints to optimize the model. The optimized model obtains the optimal corresponding keypoint set.

3.2 Principle of Feature Matching based on Rolling Guided Filter

To address the issue of viewpoint differences between satellite reference images and UAV object images, this study proposes a feature matching method based on a rolling-guided filter as shown in Figure 1. First, the keypoint response of the image is calculated using a rolling-guided filter, and the keypoints are obtained. accordingly. Second, a feature descriptor based on phase-weighted information is used to obtain the descriptors and a candidate set of corresponding keypoints is obtained through feature matching. Subsequently, the response of the candidate corresponding keypoint sets is used to compute the loss for optimizing the feature-matching network.

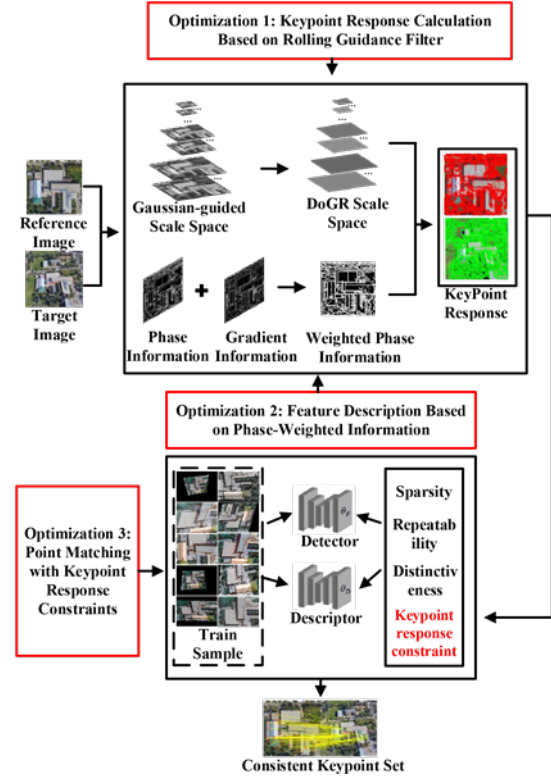


Figure 1. Property optimized feature matching framework

3.2.1 Keypoint Response Calculation based on Rolling Guided Filter

A rolling-guided filter is an edge-preserving filter that retains the edge information of an image, which is beneficial to improve the repeatability of keypoints. Therefore, in this study, the Gaussian filters represented by Equation 1 and Equation 2 are replaced with a rolling-guided filter to generate the scale pyramid. Candidate keypoints are then obtained through extreme detection in the scale pyramid, followed by subpixel localization.

(1) Rolling guided filter

The rolling-guided filter is an edge-preserving filter that can retain large-scale structural information of an image while filtering out the image noise. It consists of two parts: the removal of small-scale structures and restoration of edges.

Small-scale structures were removed using a Gaussian filter, as shown in the following equation:

$$G(p) = \frac{1}{K_p} \sum_{q \in N(p)} \exp\left(-\frac{\|p - q\|^2}{2\sigma_s^2}\right) I(q) \quad (15)$$

where $K_p = \sum_{q \in N(p)} \exp\left(-\frac{\|p - q\|^2}{2\sigma_s^2}\right)$ denotes the normalization term, I is the input image, G is the output image. p and q are the pixel coordinates in the image, σ_s is the Gaussian standard deviation, and $N(p)$ is the coordinate collection of the neighbouring pixels around p . This filter eliminates structures smaller than σ_s in scale.

Edge recovery is an iterative process. First, the input image I is smoothed using Equation 10 to obtain a

smoothed image J_1 . Then, the input image I and result of the previous iteration J_t are used in a joint bilateral filter to obtain the value of the $(t + 1)$ th iteration, denoted by J_{t+1} , as follows:

$$J^{t+1}(p) = \frac{1}{K_p} \sum_{q \in N(p)} w(p, q) \cdot I(q) \cdot J_t(q) \quad (16)$$

where $w(p, q)$ is the joint bilateral filter weight defined as:

$$w(p, q) = \exp\left(-\frac{\|I(p) - I(q)\|^2}{2\sigma_r^2} - \frac{\|J_t(p) - J_t(q)\|^2}{2\sigma_e^2}\right) \quad (17)$$

In Equation 11, $J_t(q)$ is the value of the previous iteration at pixel q , and σ_r and σ_e are the bilateral filter parameters controlling the range and domain filter standard deviations, respectively. The joint bilateral filter weight $w(p, q)$ considers the intensity similarity between p and q in the input image I , as well as the similarity between $J_t(p)$ and $J_t(q)$ in the previous iteration result J_t . This iterative process helps recover the edges in the image while preserving large-scale structures.

(2) Feature detection based on rolling guided filter

In this section, we introduce the rolling-guided filter for feature detection. The feature detection process based on the rolling-guided filter involves four steps establishing a difference in Gaussian rolling (DoGR) guided pyramid, spatial extremum detection, subpixel localization, and elimination of edge responses.

1) Establishing DoGR guided pyramid

Instead of the traditional approach, we chose a rolling guided filter to establish the difference in the Gaussian rolling (DoGR) guided pyramid. First, we set the number of octaves in the scale-space pyramid to O and the number of layers to S . The scale at any position in the scale space is given by:

$$\sigma(i, j) = k^{i-1} \sigma_0 2^{j-1} \quad (18)$$

where σ_0 is the initial scale and $k = \frac{1}{2^c}$ is the scale transfer coefficient with $S = c+3$, $i \in [1, O]$, and $j \in [1, S]$. The rolling-guided filter is convolved with the input image to obtain the scale space image, which is expressed by the following equation:

$$J^{t+1}(p, \sigma(i, j)) = \frac{1}{K_p} \sum_{q \in N(p)} \exp(f_1 + f_2) I(q) \quad (19)$$

$$f_1 = -\frac{\|p - q\|^2}{2\sigma^2(i, j)} \quad (20)$$

$$f_2 = -\frac{\|J^t(p) - J^t(q)\|^2}{2\sigma_r^2} \quad (21)$$

where $K_p = \sum_{q \in N(p)} \exp(f_1 + f_2)$ is the normalization term and $I(q)$ is the input image. The scale space was constructed by repeatedly applying Equation 14 to generate a Gaussian rolling guided pyramid with O octaves and S layers. The neighbouring images in the pyramid are then used to generate the DoGR-guided pyramid.

2) Spatial extremum detection

A local extremum can be detected on the DoGR pyramid to obtain candidate keypoints, as shown in Figure 2. Each pixel in the current pyramid layer was compared with 26 neighbouring pixels, including eight neighbouring pixels at the same level, nine neighbouring pixels at the previous level, and nine neighbouring pixels at the next level. If the value of the pixel is greater than or less than the values of these 26 pixels, the pixel is then considered to be a candidate keypoint.

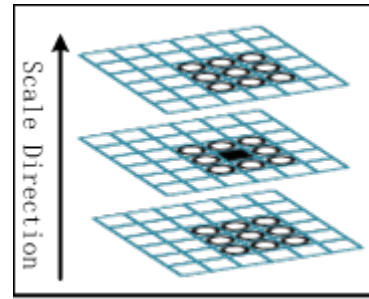


Figure 2. Spatial extremum detection

3) Subpixel localization

The extremum points obtained from the spatial extremum detection are discrete. Therefore, in this study, the precise locations and scales of the keypoints were determined by fitting a three-dimensional quadratic function. Similar to SIFT, Taylor expanded the response value of the candidate keypoint and got the $D(X)$ as follows:

$$D(X) = D + \frac{\partial D^T}{\partial X} X + \frac{1}{2} X^T \frac{\partial^2 D}{\partial X^2} X \quad (22)$$

The offset of the keypoint is calculated using the following equation:

$$\hat{X} = -\frac{\partial^2 D^{-1}}{\partial X^2} \frac{\partial D}{\partial X} \quad (23)$$

If the offsets in all three dimensions were greater than 0.5, interpolation was required to add the offsets to the keypoint, resulting in a new sampling point. Subsequently, the function value $D(\hat{X})$ at the extremum position is calculated using Equation 19, and keypoints with $|D(X)|^\wedge$ less than 0.03 are discarded.

$$D(\hat{X}) = D + \frac{1}{2} \frac{\partial D^T}{\partial X} \hat{X} \quad (24)$$

3.2.2 Feature Description based on Phase-Weighted Information

Phase congruency has the advantages of light intensity invariance or contrast invariance. Considering the significant grayscale differences between satellite reference images and UAV object images, phase-weighted information was introduced to construct feature descriptors. By computing the similarity of features, a candidate set of corresponding keypoints is generated to address the large grayscale differences.

(1) Phase congruency modelling

First, given an image I , where any pixel is denoted as $I(x, y)$. M_{so}^{even} and M_{so}^{odd} represent an evenly symmetric and odd symmetric filters of a 2D log-Gabor wavelet with scale s and orientation o . The 2D log-Gabor can be constructed as the frequency domain with a centre frequency (f_i, θ_i) , where θ_i is the orientation angle of the filter and f_i is the central radial frequency.

$$LG(f, \theta) = e^{-\frac{\log(f/f_i)}{2(\log(\sigma_{f_i}/f_i))^2}} \cdot e^{-\frac{(\theta - \theta_i)^2}{2\sigma_{\theta_i}^2}} \quad (25)$$

The relationship between the central radial frequency f_i and filter wavelength λ is given by

$$f_i = \frac{1}{\lambda} \quad (26)$$

The even and odd convolution responses $e_{so}(x, y)$ and $d_{so}(x, y)$ of image I can be obtained using Equation 22:

$$[e_{so}(x, y), d_{so}(x, y)] = [I(x, y) \cdot M_{so}^{even}, I(x, y) \cdot M_{so}^{odd}] \quad (27)$$

Next, we compute the amplitude component $A_{so}(x, y)$, phase component $\phi_{so}(x, y)$, and local energy $E(x, y)$ of the image $I(x, y)$ at scale s and orientation o :

$$A_{so}(x, y) = \sqrt{e_{so}^2(x, y) + d_{so}^2(x, y)} \quad (28)$$

$$\phi_{so}(x, y) = \text{atan2}(e_{so}(x, y), d_{so}(x, y)) \quad (29)$$

$$E(x, y) = \sqrt{\left(\sum_s \sum_o e_{so}(x, y)\right)^2 + \left(\sum_s \sum_o d_{so}(x, y)\right)^2} \quad (30)$$

Considering the influence of noise, the calculation of phase congruency is given by the following equation:

$$pc(x, y) = \frac{\sum_s \sum_o w_o(x, y) [A_{so}(x, y) \Delta \phi_{so}(x, y) - T]}{\sum_s \sum_o A_{so}(x, y) + \epsilon} \quad (31)$$

where $pc(x, y)$ represents the magnitude of the phase congruency. $w_o(x, y)$ is the weighting function, $\Delta \phi_{so}(x, y)$ is the phase deviation function, T is the noise threshold and ϵ is a small value to prevent division by zero. The $[\dots]$ operator ensures that the result is not negative and if the value inside the brackets is positive; it remain unchanged; otherwise, it is set to 0.

(2) Feature description with phase-weighted information

To achieve rotational invariance, it is necessary to assign the main orientation to the keypoints. Before constructing the feature description, the image is rotated according to the main orientation, such that the x-axis of the original image aligns with the main orientation. The following sections describe the main orientation assignment and feature description.

1) Main orientation assignment

First, the input image I is processed using Equation 26 to calculate the multi-angle phase information pc_o . Subsequently, the phase information is accumulated according to the angular direction to obtain the overall phase information pc_{all} . Next, Sobel operators are used as convolution kernels to convolve with I and pc_{all} separately and obtain the gradient magnitude gd , angle information $angle$, and phase gradient magnitude pc_{gd} .

The computation of the phase-weighted gradient information gd_{all} is given by

$$gd_{all} = \epsilon \cdot gd + (1 - \epsilon) \cdot pc_{gd} \quad (32)$$

where ϵ denotes the weighting factor, gd and gd_{pc} are calculated as follows:

$$gd = \sqrt{dx^2 + dy^2}, pc_{gd} = \sqrt{dx_2^2 + dy_2^2} \quad (33)$$

For each keypoint, a square region with a side length of 6σ and centered at the keypoint is taken. A gradient histogram was computed within this region, and the peak value was selected as the main orientation. In addition, the directions with histogram values greater than 80% of the main orientation were considered as secondary orientations.

2) Feature description

The feature description method proposed in this study is illustrated in Figure 3. Angle information and phase weighted gradient information gd_{all} of the image are computed using Equation 27. A square region with a side length of N_h centered at each keypoint was extracted, where N_h was obtained as follows:

$$N_h = \left\lfloor \min \left(12\sigma, \frac{\min(I_w, I_h)}{3} \right) \right\rfloor \quad (34)$$

where σ is the scale size of the keypoint; I_w and I_h are the width and height of the image, respectively; and $[\dots]$ denote the floor function.

Next, the two domains were then divided into

17 approximately equally sized sub-regions. Each subregion generated an 8-dimensional histogram of gradient magnitude and orientation, resulting in a final 136-dimensional feature descriptor.

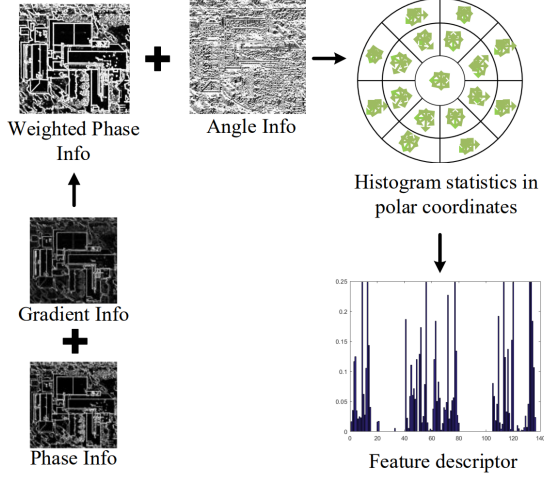


Figure 3. Illustration of the feature description process

3) Generate candidate corresponding key point set

In this section, we denote the keypoint sets obtained from the satellite reference image and UAV object image as k_1 and k_2 . The feature descriptor sets for these keypoints are denoted as d_1 and d_2 , and are generated using the phase-weighted descriptor. Euclidean distance between any two feature descriptors is given by:

$$D(d_1^i, d_2^j) = \sqrt{\sum_{n=1}^{136} (r_n^i - t_n^j)^2} \quad (35)$$

where $d_1^i = (r_1^i, r_2^i, \dots, r_{136}^i)$ for $i \in [1, 2, \dots, N_1]$, and N_1 being the number of keypoints in the satellite reference image and $d_2^j = (t_1^j, t_2^j, \dots, t_{136}^j)$ for $j \in [1, 2, \dots, N_2]$, with N_2 being the number of keypoints in a UAV object image. Feature matching was then performed based on the nearest-

neighbour criterion to obtain a candidate corresponding keypoint set, which was used to generate the network loss based on the property-optimal.

3.2.3 Feature Matching based on Property Optimization with Keypoint Response Constrained

(1) Overall process of optimized feature matching method

Traditional feature matching methods lack a connection between keypoint detection and feature description, which leads to suboptimal performance for crossview object localization in UAVs. We proposed a feature matching method based on property optimization with keypoint response constrained as shown in Figure 4.

First, images without keypoint labels are input into the keypoint detector and feature descriptor. Second, the keypoint response calculated by the rolling-guided filter and feature descriptors based on phase weighting can be used to obtain candidate corresponding keypoints. Subsequently, a loss function is constructed based on the keypoint response to jointly optimize sparsity, repeatability, and distinctiveness.

As mentioned in Section 3, our optimization is based on Equation 8. Assuming the corresponding candidate keypoint response is denoted by I_r , the loss function for optimizing the feature-matching network is given by

$$L_{all} = L_m(X, Y) + L_o(X, I_r) \quad (36)$$

$$s.t. \arg\max_{\theta_r, \theta_o} \prod_j P_s[j] P_r[j] P_g[j]$$

where X denotes the convolutional image, and Y denotes the original image. L_m is the mean-squared error loss function, and L_o is the cross-entropy loss function, which is defined as:

$$L_m(X, Y) = (X, Y)^2 \quad (37)$$

$$L_o(X, I_r) = - \sum_{h=1, w=1}^{H, W} X(h, w) \log I_r(h, w) \quad (38)$$

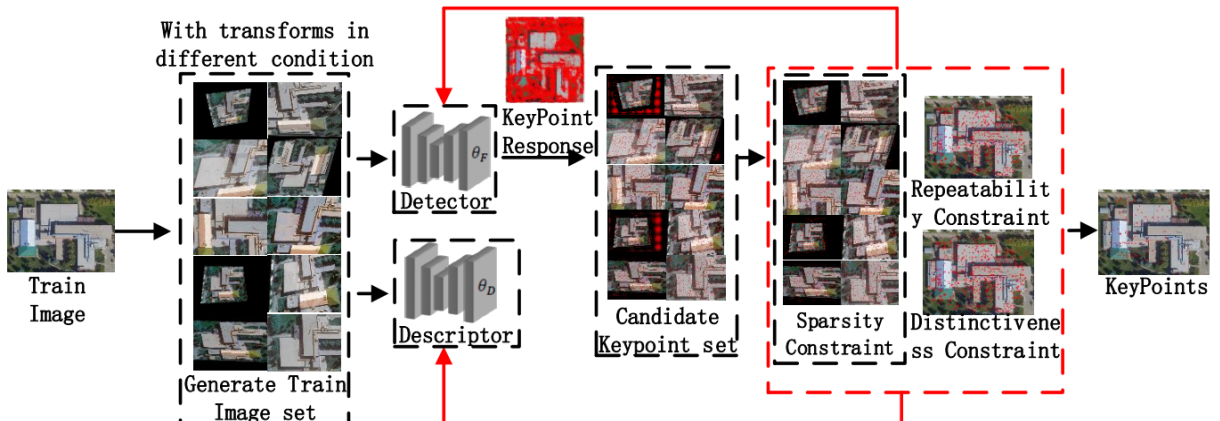


Figure 4. Illustration of optimizing feature matching method

where h and w denote the coordinates.

Finally, a keypoint detector and feature descriptor were used to compute the keypoints and feature descriptors of the images. Similar to the SIFT feature matching algorithm, the corresponding keypoint sets are obtained through nearest-neighbour matching.

4 Experimental Analysis

In the following sections, we introduce the experimental dataset, evaluation metrics, parameter analysis of the attribute optimization for keypoint constraints, comparison of keypoint matching before and after optimization, and a comparison with existing methods. The experiments described in this chapter were conducted on a PC with 16GB memory, AMD R7-5800H processor, and NVIDIA GeForce RTX 3070 graphics card. The experimental environment consisted of Python 3.8, PyTorch, and MATLAB R2016b.

4.1 Experimental Dataset and Evaluation Metrics

To validate the performance of the proposed algorithm, experiments were conducted using the University-1652 dataset released by the University of Technology Sydney (UTS). This dataset includes 1652 buildings from 72 universities worldwide and contains data from two platforms, including satellites, synthetic UAV. In this study, the first 100 classes of satellite images (one image per class) and their corresponding 100 classes of UAV images (5 images per class) were selected from the University-1652 dataset. Figure 5 shows some sample images from the dataset, which exhibit variations in scale, rotation, object types, and viewpoints.

In this study, the repeatability rate (RPR), recall rate (RR), accuracy rate (AR), and quantity rate (QR) were chosen as evaluation metrics for the experiments. Meanwhile, time efficiency (TE) is used to evaluate the running speed of different feature matching methods,

and is defined as the sum of the time taken for feature detection, feature description, and feature matching as the overall running time.

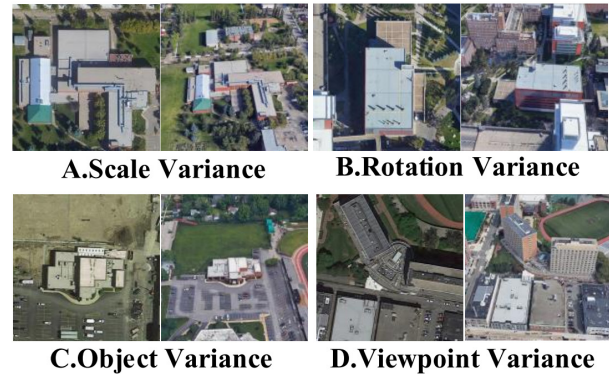


Figure 5. Sample images from the university-1652 dataset

4.2 Parameter Analysis of Property Optimization Feature Matching Method with Keypoint Constraints

In this section, we will investigate the performance of the scale factor of the rolling-guided filter, the wavelength parameter of the log-Gabor filter, and weighting factor of the phase-weighted information.

4.2.1 Analysis of Scale Factor of Rolling-guided Filter

In this section, we keep the scale factor fixed at the optimal value of 0.2 and analyse the effect of wavelength λ using a Log-Gabor filter on feature matching. A Log-Gabor filter with minimum wavelengths of 3, 6, and 9 nm was used to extract the phase congruency information from the images, as shown in Figure 6. As the wavelength increased, the details of the phase congruency information decreased, leaving only the larger object edge information. Therefore, setting the wavelength λ in the Log-Gabor filter can affect the phase congruency extraction from the image and feature matching performance.

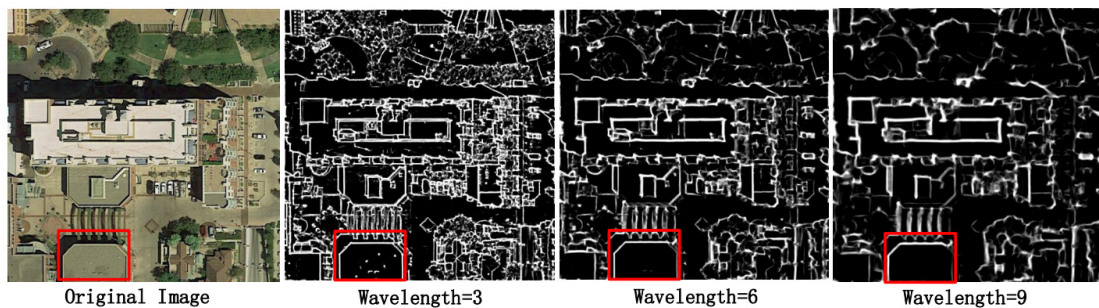


Figure 6. Phase congruency of images at different wavelengths

Figure 7 shows the results of processing using the rolling guided filter with different values of the scale factor $\sigma_r = 0.05, 0.1, 0.15, 0.2$. As the scale factor σ_r increased, the blurring of the edge areas in the image became heavier. When $\sigma_r = 0.05$, the matching accuracy (AR) was 96.36%, indicating a good matching performance. However, as σ_r increases, AR decreases to below 95%. The highest AR

of 98.44% is achieved when the $\sigma_r = 0.2$. When $\sigma_r = 0.25$, the AR drops sharply to 76.36%, indicating a significant decrease in matching performance.

Based on the above analysis, it can be concluded that the value of 0.2 for σ_r in the rolling-guided filter yields the best matching performance, with a matching accuracy exceeding 95%.

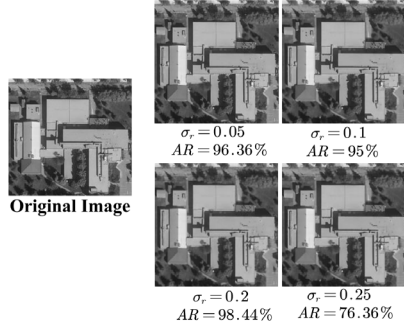


Figure 7. Variations on images with different scale

4.2.2 Analysis of Wavelength Parameter Related to Log-gabor Filter

The feature-matching comparative experiment was conducted using Log-Gabor filters of wavelengths ranging from 3 to 9, and the results are shown in Figure 8. The matching accuracy (AR) for different wavelengths are as follows: 88.06%, 92.44%, 83.83%, 91.02%, 83.93%, 91.70%, and 87.80%. After wavelength 9, AR meet a reduce. Regardless of the wavelength, the AR was higher than 80%. When the wavelength is set to 4, 6, or 8, the AR can exceed 90%. Moreover, when the wavelength was set to 4, the matching accuracy reached its highest value of 92.44.

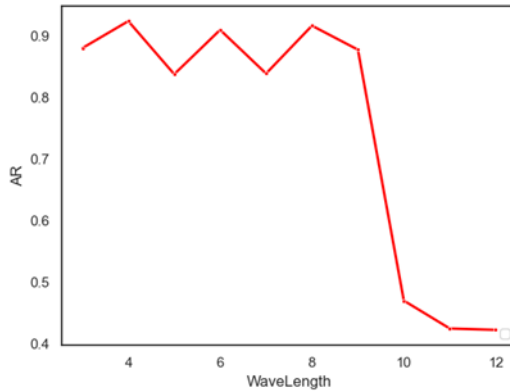


Figure 8. Matching accuracy at different wavelengths

Therefore, we set the wavelength parameter to 4 can reduce noise interference, extract better phase congruency information from the image, and improves feature matching performance.

4.2.3 Analysis of Weighting Factor for Phase-weighted Information

The weighting factor ϵ is used to balance the phase and gradient information. In this section, the scale factor of the rolling-guided filter is set to 0.2, wavelength of the Log-Gabor filter is set to 4, and ϵ is set from 0.1 to 0.9 for conducting comparative experiments on feature matching.

The experimental results are presented in Figure 9. When ϵ was set from 0.1 to 0.9, with a step length of 0.1, the matching accuracy (AR) was 94.36%, 89.85%, 96%, 67.2%, 66.66%, 63.15%, 94.15%, 60%, and 74.07%, respectively. Overall, when ϵ is in the range of 0.1–0.3, the AR exceeds 89% and shows good matching performance,

with the highest AR achieved at $\epsilon = 0.3$. Subsequently, in the range of 0.4 to 0.6, AR remains between 60% and 70% with a significant decline in matching performance. When $\epsilon = 0.7$, the matching accuracy increases to 94.91%.

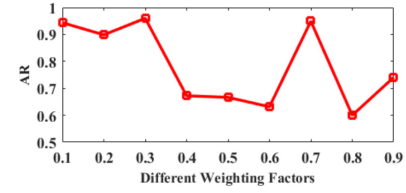


Figure 9. Matching accuracy with different weight factors

Then, in the range of 0.8 to 0.9, the AR is consistently below 75%. When $\epsilon = 0.8$, the AR reaches the lowest value of 60%. Based on the above analysis, ϵ values of 0.1, 0.2, 0.3, and 0.7 yield a good matching performance, with the optimal performance achieved at $\epsilon = 0.3$.

4.3 Comparison and Analysis of Feature Matching Methods Before and After Optimization

The method proposed in this section is a feature-matching method using a rolling-guided filter. Through experimental analysis, we set the weight factor of the rolling filter to 0.2, Log-Gabor filter wavelength to 4, and phase information weight factor to 0.3. The results are then compared and analysed with SIFT [27] and POP-Net [44] as listed in Table 1.

Table 1. Comparison of methods before and after optimization

	RPR	RR	AR	QR
SIFT	0.7512	0.0020	0.1980	0.0115
POP-Net	0.5825	0.0146	0.2317	0.0033
Optimized	0.9801	0.0324	0.4460	0.2504

According to the previously mentioned criteria, the repeatability rate (RPR) of SIFT, POP-Net, and the proposed method were 0.7512, 0.5825, and 0.9801,

respectively. Compared to SIFT and POP-Net, the proposed method shows improvements of 22.89% and 39.76% in RPR, respectively. This indicates that the optimized method can detect more repeatable keypoints in the same scene under different environmental conditions.

The recall rates (RR) of SIFT, POP-Net, and the proposed method were 0.0020, 0.0146, and 0.0324, respectively. The proposed method achieved the highest RR, indicating that the keypoints detected by the optimized method were more distinctive and suitable for matching.

Furthermore, the accuracy rates (AR) of SIFT, POP-Net, and the proposed method were 0.198, 0.2317, and 0.446, respectively. The AR of the optimized feature matching method was the highest, showing improvements of 24.8% and 21.43% compared with SIFT and POP-Net, respectively.

Finally, the quantization rate (QR) of SIFT, POP-Net, and the proposed method were 0.0115, 0.0033, and

0.2504, respectively. The QR of the optimized feature matching method was the highest, showing improvements of 23.89% and 24.71% compared with SIFT and POP-Net, respectively.

Based on the above analysis, the proposed method performed better than the pre-optimized method for all evaluation metrics. Therefore, the proposed method is more suitable for solving UAV cross-view feature matching problems.

4.4 Comparison and Analysis with Existing Methods

In this section, the proposed method is compared with several mainstream algorithms including SIFT [27], SURF [45], ORB [46], ASIFT, AKAZE [47], Superpoint [48], SuperGlue [49], DISK [50], TILDE [51], and POP-Net [44]. The experimental results are presented in Table 2.

Among all algorithms, ORB has the fastest matching speed, but its matching accuracy (AR) was only 0.1690, which does not effectively solve the problem of matching of satellite reference images and UAV object images. SIFT, SURF, AKAZE, and ASIFT have AR values of 0.1980, 0.2240, 0.2390, and 0.4290, respectively, all of which were higher than that of ORB, and SURF, AKAZE, and ASIFT have higher matching precision than SIFT. Among them, ASIFT performs well in large image viewpoint differences, with RPR of 0.9665, RR of 0.0046, and QR of 0.1608, but the time efficiency was significantly reduced, with an average matching time of 14.6s for two 512*512 pixel images. In deep learning-based feature matching methods, TILDE, Superpoint, DISK, POP-Net, and SuperGlue have AR values of 0.0038, 0.2722, 0.1860, 0.2317, and 0.2775, respectively. SuperGlue has the highest AR and RR values, whereas the performances of Superpoint and POP-Net were slightly lower than those of SuperGlue. DISK has the highest repeatability rate (RPR) and quantization rate (QR) among these methods, with an RPR of 0.9075 and a QR of 0.0124. The feature matching method proposed in this study had RPR, RR, AR, and QR values of 0.9801, 0.0324, 0.446, and 0.2504, respectively, with the highest RPR, AR, and QR values overall. Compared to the other methods, the proposed method improved the RPR, AR, and QR values by 32.25%, 22.3%, and 22.93%, respectively. The experimental data indicate that the proposed method can better address the issue of viewpoint differences in cross view point feature matching, thereby laying the foundation for further research on feature matching.

Figure 10 and Figure 11 demonstrate the performance of different feature matching algorithms for building and stadium. As shown in Figure 10, deep learning-based methods, such as Superpoint, SuperGlue, DISK, TILDE, and POP-Net outperform traditional feature-matching methods, except for ASIFT. In Figure 11, where there was a significant variation in the viewpoint between the two images. ORB, SURF, AKAZE, DISK, SuperGlue, Superpoint, and POP-Net failed to provide accurate matches in these two images, whereas the proposed method in this study maintains a good matching performance and effectively addresses the issue of viewpoint differences in feature matching.

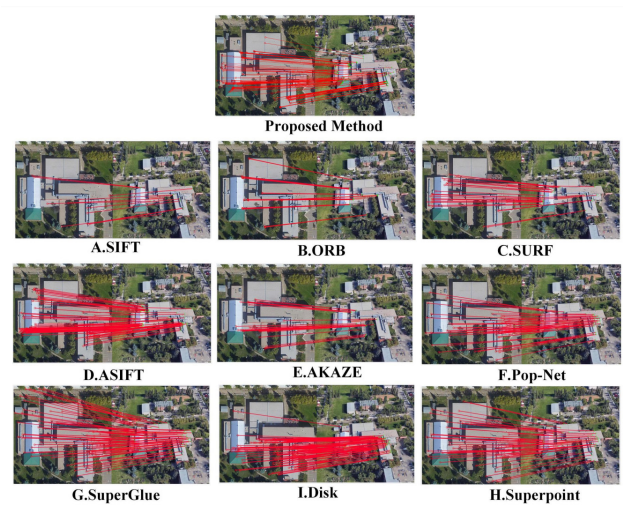


Figure 10. Matching results of building images

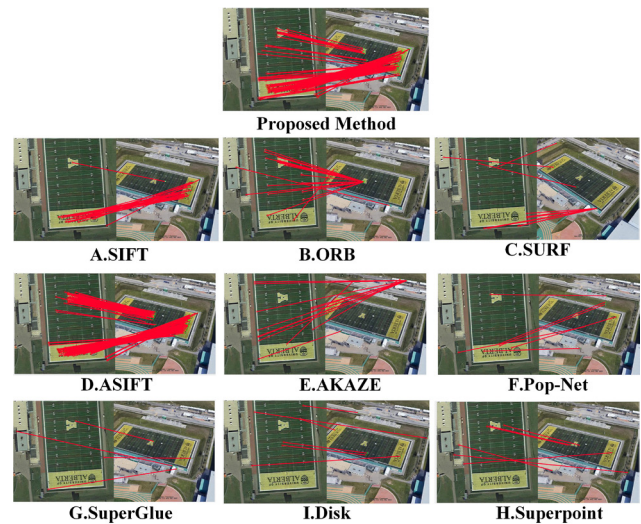


Figure 11. Matching results of stadium images

5 Conclusion

In this study, a rolling-guided filter-optimized feature matching method was proposed to address the issue of viewpoint differences between satellite reference and UAV target images. Phase congruency was incorporated in the feature description step to handle the grayscale differences between the two images using phase-weighted information. Subsequently, a candidate set of corresponding keypoints was obtained using the previous methods, and the correspondences of these keypoints were used to construct a network loss model for the joint optimization of sparsity, repeatability, and discriminability to obtain keypoint detector and feature descriptor. Finally, the trained models were used to extract keypoints and feature descriptors from two images, and the corresponding keypoint set was obtained through nearest-neighbour matching. To validate the effectiveness of the proposed method, experiments were conducted using a University-1652 dataset. The proposed method achieved an average improvement of 32.25% in the repeatability rate (RPR), 22.3% in the

accuracy rate (AR), compared with the baseline methods. Therefore, the proposed method is effective in addressing

viewpoint differences between satellite reference images and UAV object images.

Table 2. Comparison with existing feature matching method

	RPR	RR	AR	QR	Time(s)
SIFT	0.7512	0.0020	0.1980	0.0115	1.20
SURF	0.8799	0.0020	0.2240	0.0112	0.93
ORB	0.3811	0.0242	0.1690	0.0019	0.63
AKAZE	0.7032	0.0067	0.2390	0.0042	1.39
ASIFT	0.9665	0.0046	0.4290	0.1608	14.60
TILDE	0.2794	0.0015	0.0038	0.0019	2.56
DISK	0.9075	0.0033	0.1860	0.0124	1.50
Superpoint	0.5392	0.0245	0.2722	0.0021	0.72
SuperGlue	0.5849	0.0649	0.2775	0.0026	2.20
POP-Net	0.5831	0.0146	0.2317	0.0033	0.67
Proposed	0.9801	0.0324	0.4460	0.2504	0.38

Acknowledgment

The authors would like to thank the anonymous reviewers for their valuable comments and suggestions on the paper. This work is supported by the National Natural Science Foundation of China (6227010741), the Natural Science Foundation of Heilongjiang (LH2022E114), and National Natural Science Foundation Training Project of Jiamusi University (JMSUGPZR2022-016), Doctoral Program of Jiamusi University (JMSUBZ2022-13).

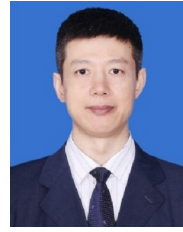
References

- [1] J. Sun, B. Li, Y. Jiang, W. Chin-Yung, A camera-based target detection and positioning UAV system for search and rescue (SAR) purposes, *Sensors*, Vol. 16, No. 11, Article No. 1778, November, 2016.
- [2] N. Kerle, F. Nex, M. Gerke, D. Duarte, A. Vetrivel, UAV-based structural damage mapping: A review, *ISPRS international journal of geo-information*, Vol. 9, No. 1, Article No. 14, January, 2020.
- [3] W. H. Maes, K. Steppe, Perspectives for remote sensing with unmanned aerial vehicles in precision agriculture, *Trends in plant science*, Vol. 24, No. 2, pp. 152–164, February, 2019.
- [4] Z. X. Li, Z. K. Yan, X. Zhang, Discussion on Key Technology and Application of UAV Surveying Data Processing, *Bulletin of Surveying and Mapping*, No. S1, pp. 36–40, January, 2017.
- [5] L. D. P. Pugliese, F. Guerriero, Last-mile deliveries by using drones and classical vehicles, *International Conference on Optimization and Decision Science: Methodologies and Applications*, Springer, Sorrento, Italy, 2017, pp. 557–565.
- [6] L. Liu, H. Li, Lending orientation to neural networks for cross-view geo-localization, *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, Long Beach, CA, USA, 2019, pp. 5617–5626.
- [7] S. Hu, M. Feng, R. M. H. Nguyen, G. H. Lee, Cvm-net: Cross-view matching network for image-based ground-to-aerial geo-localization, *IEEE Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, USA, 2018, pp. 7258–7267.
- [8] S. Zhu, T. Yang, C. Chen, Vigor: Cross-view image geo-localization beyond one-to-one retrieval, *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Nashville, TN, USA, 2021, pp. 3639–3648.
- [9] Z. Zheng, Y. Wei, Y. Yang, University-1652: A multi-view multi-source benchmark for drone-based geo-localization, *28th ACM international conference on Multimedia*, Seattle, WA, USA, 2020, pp. 1395–1403.
- [10] Q. P. Zhang, Y. Cao, Research on Three-Dimensional Reconstruction Algorithm of Weak Textured Objects in Indoor Scenes, *Laser & Optoelectronics Progress*, Vol. 58, No. 8, pp. 0810017-1–0810017-7, April, 2021.
- [11] H. H. Duan, Y. Y. Liu, H. X. Huang, Z. T. Wang, H. W. Zhao, Image Stitching Algorithm for Drones Based on SURF-GHT, *IOP Conference Series: Materials Science and Engineering*, Vol. 569, No. 5, Article No. 052025, August, 2019.
- [12] Y. G. Li, Z. K. Zhou, L. Bai, Large-scale Automatic Driving Scene Reconstruction Based on Binocular Image, *Computer Science*, Vol. 46, No. 11A, pp. 251–254, November, 2019.
- [13] T. Y. Liu, D. Jia, S. L. Luo, K. Wang, Construction and Matching of Thermal Feature Descriptor for Image Retrieval, *Journal of Frontiers of Computer Science and Technology*, Vol. 17, No. 1, pp. 166–178, January, 2023.
- [14] S. Zou, P. Tang, C. M. Hu, X. J. Shan, Full-automatic Splicing Method of Unmanned Aerial Vehicle Images in Large Area Based on 3D Reconstruction, *Computer Engineering*, Vol. 45, No. 4, pp. 235–240, April, 2019.
- [15] S. Liu, B. Yang, Y. Wang, J. Tian, L. Yin, W. Zheng, 2D/3D Multimodal Medical Image Registration Based on Normalized Cross-Correlation, *Applied Sciences*, Vol. 12, No. 6, Article No. 2828, March, 2022.
- [16] F. W. Hu, Mutual information-enhanced digital twin promotes vision-guided robotic grasping, *Advanced Engineering Informatics*, Vol. 52, Article No. 101562, April, 2022.
- [17] D. Xiang, Y. Xie, J. Cheng, Y. Xu, H. Zhang, Y. Zheng, Optical and SAR Image Registration Based on Feature Decoupling Network, *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 60, pp. 1–13, October, 2022.

- [18] P. Wu, H. L. Xu, W. L. Song, A fast NCC image matching algorithm based on wavelet pyramid search strategy, *Journal of Harbin Engineering University*, Vol. 38, No. 5, pp. 791-796, May, 2017.
- [19] J. Yu, Y. Chen, S. Li, H. Zhang, Y. Chen, Secondary Matching Algorithm: a New Heterogeneous Image Matching Algorithm for the UAV Image and Satellite Remote Sensing Image, *IGARSS 2022-2022 IEEE International Geoscience and Remote Sensing Symposium*, Kuala Lumpur, Malaysia, 2022, pp. 3275-3278.
- [20] Y. J. Fu, K. F. Sun, J. L. Liu, A Fast Anti-rotation Scene Matching Algorithm Based on Covariance Matrix, *Journal of Projectiles, Rockets, Missiles and Guidance*, Vol. 42, No. 5, pp. 20-26, 2022.
- [21] J.-S. Pan, B. Sun, S.-C. Chu, M. Zhu, C.-S. Shieh, A Parallel Compact Gannet Optimization Algorithm for Solving Engineering Optimization Problems, *Mathematics*, Vol. 11, No. 2, Article No. 439, January, 2023.
- [22] P. Hu, J.-S. Pan, S.-C. Chu, C. Sun, Multi-surrogate assisted binary particle swarm optimization algorithm and its application for feature selection, *Applied Soft Computing*, Vol. 121, Article No. 108736, May, 2022.
- [23] P. Polewski, W. Yao, Scale invariant line-based co-registration of multimodal aerial data using L1 minimization of spatial and angular deviations, *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 152, pp. 79-93, June, 2019.
- [24] A. Hafiane, K. Palaniappan, G. Seetharaman, UAV-video registration using block-based features, *IGARSS 2008-2008 IEEE International Geoscience and Remote Sensing Symposium*, Boston, MA, USA, 2008, pp. II-1104-II-1107.
- [25] X. Wang, A. Kealy, W. Li, B. Jelfs, C. Gilliam, S. Le May, B. Moran, Toward autonomous uav localization via aerial image registration, *Electronics*, Vol. 10, No. 4, Article No. 435, February, 2021.
- [26] D. G. Lowe, Distinctive image features from scale-invariant keypoints, *International journal of computer vision*, Vol. 60, No. 2, pp. 91-110, November, 2004.
- [27] Y. Ke, R. Sukthankar, PCA-SIFT: A more distinctive representation for local image descriptors, *2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Washington, DC, USA, 2004, pp. 1-8.
- [28] I. Kovač, P. Marak, Finger vein recognition: utilization of adaptive gabor filters in the enhancement stage combined with sift/surf-based feature extraction, *Signal, Image and Video Processing*, Vol. 17, No. 3, pp. 635-641, April, 2023.
- [29] J. Gao, Z. Sun, An Improved ASIFT Image Feature Matching Algorithm Based on POS Information, *Sensors*, Vol. 22, No. 20, Article No. 7749, October, 2022.
- [30] Y. X. Niu, M. Q. Chen, H. Zhang, Fast Scene Matching Method Based on Scale Invariant Feature Transform, *Journal of Electronics and Information Technology*, Vol. 41, No. 3, pp. 626-631, March, 2019.
- [31] W. Y. Zhang, Z. Li, Y. J. Wang, UAV scene matching algorithm based on CenSurE-star feature, *Chinese Journal of Scientific Instrument*, Vol. 38, No. 2, pp. 462-470, February, 2017.
- [32] K. M. Yi, E. Trulls, V. Lepetit, P. Fua, Lift: Learned invariant feature transform, *Computer Vision-ECCV 2016: 14th European Conference*, Amsterdam, Netherlands, 2016, pp. 467-483.
- [33] Y. Yuan, W. Huang, X. Wang, H. Xu, H. Zuo, R. Su, Automated accurate registration method between UAV image and Google satellite map, *Multimedia Tools and Applications*, Vol. 79, No. 23-24, pp. 16573-16591, June, 2020.
- [34] M. H. Mughal, M. J. Khokhar, M. Shahzad, Assisting UAV localization via deep contextual image matching, *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, Vol. 14, pp. 2445-2457, January, 2021.
- [35] H. Goforth, S. Lucey, GPS-denied UAV localization using pre-existing satellite imagery, *2019 International Conference on Robotics and Automation (ICRA)*, Montreal, QC, Canada, 2019, pp. 2974-2980.
- [36] J. Kinnari, F. Verdoja, V. Kyrki, Season-invariant GNSS-denied visual localization for UAVs, *IEEE Robotics and Automation Letters*, Vol. 7, No. 4, pp. 10232-10239, October, 2022.
- [37] J. Choi, H. Myung, BRM localization: UAV localization in GNSS-denied environments based on matching of numerical map and UAV images, *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Las Vegas, NV, USA, 2020, pp. 4537-4544.
- [38] A. Nassar, K. Amer, R. ElHakim, M. ElHelw, A deep cnn-based framework for enhanced aerial imagery registration with applications to uav geolocalization, *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, Salt Lake City, UT, USA, 2018, pp. 1594-1604.
- [39] C. Xu, C. Liu, H. Li, Z. Ye, H. Sui, W. Yang, Multiview Image Matching of Optical Satellite and UAV Based on a Joint Description Neural Network, *Remote Sensing*, Vol. 14, No. 4, Article No. 838, February, 2022.
- [40] C. Harris, M. Stephens, A combined corner and edge detector, *Alvey vision conference*, Manchester, UK, 1988, pp. 1-6.
- [41] K. Mikolajczyk, C. Schmid, Scale & affine invariant interest point detectors, *International journal of computer vision*, Vol. 60, No. 1, pp. 63-86, October, 2004.
- [42] P. Kovesi, Image features from phase congruency, *Videre: Journal of computer vision research*, Vol. 1, No. 3, pp. 1-26, Summer, 1999.
- [43] X. Liu, J.-B. Li, J.-S. Pan, S. Wang, X. Lv, S. Cui, Image-matching framework based on region partitioning for target image location, *Telecommunication Systems*, Vol. 74, No. 3, pp. 269-286, July, 2020.
- [44] P. Yan, Y. Tan, Y. Tai, D. Wu, H. Luo, X. Hao, Unsupervised learning framework for interest point detection and description via properties optimization, *Pattern Recognition*, Vol. 112, pp. 1-13, April, 2021.
- [45] H. Bay, T. Tuytelaars, L. Van Gool, Surf: Speeded up robust features, *9th European Conference on Computer Vision (ECCV 2006)*, Graz, Austria, 2006, pp. 404-417.
- [46] E. Rublee, V. Rabaud, K. Konolige, G. Bradski, ORB: An efficient alternative to SIFT or SURF, *2011 International conference on computer vision*, Barcelona, Spain, 2011, pp. 2564-2571.
- [47] P. F. Alcantarilla, J. Nuevo, A. Bartoli, Fast explicit diffusion for accelerated features in nonlinear scale spaces, *British Machine Vision Conference*, Bristol, UK, 2013, pp. 1-12.
- [48] D. DeTone, T. Malisiewicz, A. Rabinovich, Superpoint: Self-supervised interest point detection and description, *IEEE conference on computer vision and pattern recognition workshops*, Salt Lake City, UT, USA, 2018, pp. 337-349.
- [49] P.-E. Sarlin, D. DeTone, T. Malisiewicz, A. Rabinovich,

Superglue: Learning feature matching with graph neural networks, *IEEE/CVF conference on computer vision and pattern recognition*, Seattle, WA, USA, 2020, pp. 4937-4946.

- [50] M. Tyszkiewicz, P. Fua, E. Trulls, DISK: Learning local features with policy gradient, *NIPS'20: Proceedings of the 34th International Conference on Neural Information Processing Systems*, Vancouver, BC, Canada, 2020, pp. 14254-14265.
- [51] Y. Verdie, K. Yi, P. Fua, V. Lepetit, Tilde: A temporally invariant learned detector, *IEEE conference on computer vision and pattern recognition*, Boston, MA, USA, 2015, pp. 5279-5288.



Huaqi Zhao received Bachelor's Degree from school of Industrial Automation, Harbin Engineering University, in 1999, Master's Degree from the school of Mechanical and Electrical Engineering, Northeast Forestry University, in 2005, and Doctor's Degree from school of Mechanical and Electrical Engineering, Northeast Forestry University, in 2009. His research focuses on pattern recognition and artificial intelligence.

Biographies



Xiaomin Liu received Bachelor's Degree from school of Information and Electronic Technology, Jiamusi University, in 2003, Master's Degree from the school of computer science, Heilongjiang University, in 2009, and Doctor's Degree from School of Electronics and Information Engineering, Harbin of Institution Technology, in 2022. Her research focuses on image process, pattern recognition, image matching and artificial intelligence.



Yizhao Yuan received Bachelor's Degree of computer science in School of Information Engineering, China Jiliang University in 2021. He is studying master's degree in Jiamusi University. His research focuses on image processing.



Donghua Yuan received the Bachelor's Degree in Shaoguan University in 2020, the Master's degree in Jiamusi University in Computer Applied Technology in 2023. His research focuses on image processing.



Jeng-Shyang Pan received the B. S. degree in Electronic Engineering from the National Taiwan University of Science and Technology in 1986, the M. S. degree in Communication Engineering from the National Chiao Tung University, Taiwan in 1988, and the Ph.D. degree in Electrical Engineering from the University of Edinburgh, U.K. in 1996.